DOCUMENT RESUME

ED 233 314 CS 007 245

AUTHOR Anderson, Jonathan
TITLE Cohesion, Comprehending and Comprehensibility.
PUB DATE May 83
NOTE 16p.; Paper presented at the Annual Meeting of the International Reading Association (28th, Anaheim, CA, May 2-6, 1983).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Coherence; *Cohesion (Written Composition); *Difficulty Level; Elementary Secondary Education; *Evaluation Methods; Reading Materials; *Reading Material Selection; Reading Research
IDENTIFIERS *Linguistic Analysis; *Text Coherence; Text Structure

ABSTRACT

At a workshop exploring practical methods that elementary and secondary teachers may use to analyze texts, three parallel texts were subjected to three forms of analysis: subjective analysis, examination of cohesive ties, and linguistic analysis of selected text variables such as word and sentence length. The subjective ranking of the passages' difficulty illustrated the varied factors on which individual judgments are based. Calculation of the percentage of grammatical and lexical ties per word and per sentence failed to correspond to workshop participants' subjective judgments of difficulty, suggesting that the connection between cohesion and comprehensibility needs to be investigated further. Rix scores--obtained for each passage by counting the number of long words, dividing by the number of sentences, and multiplying by 100--corresponded roughly to the subjective ratings. While providing an objective measure, the Rix scale's equation of sentence and word length with sentence and word complexity is questionable. Although these measures are indirect, they may nevertheless provide teachers with help in critically evaluating classroom texts. (MM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED233314

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

COHESION, COMPREHENDING AND COMPREHENSIBILITY

Jonathan Anderson
Flinders University of South Australia

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Jonathan Anderson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented to the 28th Annual Convention of the International Reading Association, Anaheim, California, May 1983.

CS 007 245

The purpose of this workshop session was to examine practical procedures that teachers, both primary and secondary, may use to analyse texts. Two commonly employed procedures for assessing either reading comprehension or the comprehensibility of texts focus on word factors (tapping lexical difficulty) and sentence factors (tapping syntactic complexity). However, as teachers know well, while many a reader, native speaker and second-language learner, may be familiar with the lexical items and syntactic structures within a text, difficulties are often still experienced with comprehending the text as a whole. More recently, researchers have provided a framework for looking at factors beyond the sentence, termed cohesive ties, that link text together providing continuity and texture.

The workshop provided opportunities to analyse texts in terms of the cohesive properties and other linguistic factors which may serve to anticipate the ease or difficulty of reading them.

Subjective analysis

Three parallel texts (see next page) were distributed and participants were invited to rank these in order of difficulty as they might appear to young readers. There was some discussion on the group rank orderings in an effort to discover what aspects of texts teachers focused on in making their judgments. Although there was far from perfect agreement, the majority view placed Passage C as most difficult and Passage B as easiest. The discussion illustrated the subjective nature of global judgment as well as the fact that individual judgments were often based on a variety of different factors. The major purpose of the exercise was not to disparage subjective analyses but to begin to focus attention on the major characteristics of text.

Three parallel texts to be placed on next page following first line of above paragraph

---

It was readily apparent that the register was constant across texts. In other words, not only was the story the same but the storyteller's purpose was similar in each case, as was the style. In Halliday's terms,

the field, mode and tenor were similar for all three texts. This meant, then, that if the texts did indeed differ in difficulty then such difficulty could be attributed essentially to text variables rather than factors of style or genre.

Cohesive ties

Attention was directed first to cohesive ties, the semantic links which according to Halliday and Hasan (1976) "make a text a text" (p.13). Another writer has said of Cohesive ties:

> ... the cohesive ties which bind or bridge sentences semantically enable the reader to establish coherence in text and therefore account for a large portion of ... comprehensibility in written discourse (Moe 1979, p.16).

To illustrate how linguistic ties quite literally hold text together, like "a kind of linguistic mortar" as Tierney and Mosenthal (1980) expressed it, a tracing of the ties in the nursery rhyme of Old Mother Hubbard was made:

Old Mother Hubbard rhyme about here

Halliday and Hasan (1976) in their pioneering book Cohesion in English identified five major categories of cohesion. These categories - reference, substitution, ellipsis, conjunction and lexical cohesion - were in turn each sub-divided into two or more sub-categories as illustrated in the Figure below.

Halliday and Hasan's categories of cohesion about here

For those unfamiliar with the concept of cohesion the following description from Anderson (1982) illustrates the major types and sub-categories:

1. Reference

Reference, as the name suggests, requires reference elsewhere in the text for interpretation of one item in the tie. Halliday and Hasan identify three sub-categories of reference,

a) Personal reference

e.g. 1. Joan went to town.

2. She bought herself a hat.

(She and herself refer to Joan).

This sub-category includes the personal pronouns, possessive adjectives and possessive pronouns (I, me, mine, my, he, him, his, it, its, one, one's etc.)

b) Demonstrative reference

e.g. 3. Joan went to town.

4. There she bought a hat.

5. The hat was a delicate shade of pink.

(There refers to town and the to the same hat as in 4).

This sub-category includes the demonstratives (this, that, these, those, here, there, now, then, the).

c) Comparative reference

e.g. 6. Joan bought a hat.

7. It is the same as mine.

(same refers to hat).

This sub-category includes comparative adjectives and adverbs (some, identical, similarly, other, better, more, so etc.)

2. Substitution

In substitution one item (usually one, do or so) substitutes for another. Halliday and Hasan describe three sub-categories of substitution depending on whether the substitute functions as a noun, verb or clause.

a) Nominal substitution

e.g. 8. Robert got a new bicycle for Christmas.

9. His old one was broken.

(one substitutes for bicycle)

The words occurring as nominal substitutes are: one, ones; same

b) Verbal substitution

e.g. 10. Robert got a new bicycle.

11. I did too.

(did substitutes for got a new bicycle)

The verb do is the only verb identified by Halliday and Hasan that occurs as a verbal substitute.

c) Clausal substitution

e.g. 12. Is Robert getting a new bicycle?

13. I hope so.

(so substitutes for Robert is getting a new bicycle).

Along with so, not can substitute for a clause.

3. Ellipsis

Ellipsis is rather like substitution and is in fact described by Halliday and Hasan as "substitution by zero". As in substitution, three sub-categories are delineated depending on whether the presumed item functions as a noun, verb or clause.

a) Nominal ellipsis

e.g. 14. Mary had a chocolate milkshake.

15. Jean has a vanilla.

(a vanilla milkshake is presumed).

b) Verbal ellipsis

e.g. 16. Did Mary do away?

17. She may have.

(gone away is presumed).

c) Clausal ellipsis

e.g. 18. Mary was having a milkshake.

19. Who was?

(having a milkshake is presumed).

4. Conjunction

Conjunctions serve to connect one part of a text with another and thus form cohesive ties. Halliday and Hasan (1976) describe such relations as indirect in that conjunctions "express certain meanings which presuppose the presence of other components in the discourse" (p.226). They identify four sub-categories.

(a) Additive (and, also, nor, in addition, by the way etc.)

(b) Adversative (but, yet, though, however, on the other hand etc.)

(c) Causal (so, hence, as a result, therefore etc.)

These four sub-categories are illustrated in the following passage (pp.238-239).

14. For the whole day he climbed up the steep mountain-side, almost without stopping.

15. And in all this time he met no one. Additive

16. Yet he was hardly aware of being tired. Adversative

17. So by night-time the valley was far below him. Causal

18. Then, as dusk fell, he sat down to rest. Temporal

5. Lexical cohesion

Reference, substitution, ellipsis and conjunction are types of cohesion that are "realised" through the grammar. A fifth type of cohesion is realised through the vocabulary. Halliday and Hasan term this lexical cohesion. They identify two sub-categories.

a) Reiteration

i.e. the repetition of the same item, or the use of a synonym (or near synonym), a superordinate, or a general noun. For example:

19. I turned to the ascent of the peak.	
20. The ascent is perfectly easy.	Repetition
21. The climb is perfectly easy.	Synonym
22. The task is perfectly easy.	Superordinate
23. The thing is perfectly easy.	General noun

b) Collocation

i.e. "the association of items that regularly co-occur" (Halliday and Hasan, p.285). For example, night may be associated with day, book with newspaper, private with sergeant, curtain with window...etc.

Analysis of cohesive ties

As part of the workshop the cohesive ties in each of the three parallel texts were identified. Just the overall results are presented here. Thus Table 1 presents, for each of the three texts, the total pattern of cohesive ties, sub-divided into grammatical ties (reference, substitution and ellipsis) and lexical ties. Following Hasan (1980), conjunctions were omitted from this part of the analysis.

There is no reason, of course, why the total number of ties, or for that matter the number of grammatical or lexical ties, should show any variation across texts since these counts would be a function of text length. Accordingly, the individual counts were first expressed as percentages. Examination of Table 1 shows that when this was done the pattern in Passages B and C appear very similar with the breakdown of grammatical to lexical ties being approximately 60:40 whereas in Passage A it is nearer 70:30. In other words, Passage A is the one that is different though whether this is a significant difference or not is unclear.

Another way of taking text length into account is to express number of ties as a function of number of words or number of sentences. When this was done, as is also seen in Table 1, Passages A and C seem remarkably similar with Passage B being the odd one out. The metric, ties per word, is a measure of the tie density. Passage B on this metric has the greatest tie density. Is tie density related to reading difficulty? It could perhaps be argued that the greater number of ties or connections readers need to make, then the more complex the reading task and hence tie density is a measure of reading difficulty. On this reasoning, Passage B would be the most difficulty of the three passages to read and the remaining two passages would be approximately equally difficult. On the other hand, perhaps it could be argued that the greater the rate of cohesion the more tightly knit is the text and maybe this is really an indicator of reading ease. On this line of argument, Passage B would be the easiest to read.

The measure ties per sentence has perhaps greater grounds for being related to reading difficulty for practically every readability formula is based on the notion that sentence length is a measure of syntactic complexity. The argument here would be that shorter sentences contain fewer words and thus would be expected to have a lower rate of ties per sentence. Certainly, the rate of ties per sentence for Passage B of 4.00 appears to be considerably lower than the rates for either of the other two passages (7.00 and 7.75 respectively). This line of reasoning would suggest that Passage B is the easiest to read with little difference in reading difficulty between Passages A and C.

Table 1 about here

Analysis of other selected text variables

The next stage was to focus attention on more traditional text variables in an effort to see whether analysis of these variables would shed further light on the tentative results of the cohesion analysis or on the group subjective ratings.

A count was therefore made of the number of words, long words (defined as containing more than six letters), syllables, monosyllables and sentences since these are the basic measures included in most readability formulae (see Table 2). Formulae incorporating percentage of syllables, for example, and sentence length are those of Flesch and Fry. Possibly more useful than either of these is a new formula Lix, and a super-fast version Rix (Anderson 1983). Lix is defined simply as follows:

Lix = word length + sentence length
where word length is percentage of long words
and sentence length is average length of sentence in words

Table 2 includes these two calculations for the three texts and summing gives Lix scores of 30.1, 20.5 and 30.5 respectively for Passages A, B and C. Passage B on the Lix criterion is thus seen to be the easiest and Passages A and C of the approximately equal difficulty.

Table 2 about here

Before trying to interpret these Lix scores, let us note one of the major advantages of Lix over most other readability estimates. By examining the components of Lix, the relative contributions of sentence and word factors to estimated reading difficulty are readily apparent. Thus for Passage A (and somewhat similarly for Passage C) the sentence factor (20.2) contributes approximately twice as much to reading difficulty as the word factor (9.9). Furthermore, these factor weights are themselves readily interpretable. A sentence factor of 20.2, for instance, indicates an average sentence length of about 20 words; while a word factor of 9.9 indicates that about 10 per cent of the words, that is an average of one word in ten, are long words. Passage B, by contrast, has the word factor as the major contributor to reading difficulty (approximately 12/20 or 60 per cent) compared with the contribution of the sentence factor (approximately 8/20 or 40 per cent). Again, these factor weights are readily interpretable.

While Lix is very easy to calculate, its close relative Rix is even easier. The full directions and interpreter appear on the next page. All that is required is to count the number of long words, divide by the number

of sentences and multiply by 100. The interpreter allows the user to enter the Rix score in the Table and read off the equivalent grade level. Thus for Passage A:

Rix = (10/5) x 100
= 200 (equivalent to Grade 6)

Passage B is estimated at Grade 4 level and Passage C at Grade 6 level.

Estimating Reading Difficulty with Rix on next full page following second sentence in above paragraph

---

Discussion

In the analysis of any particular text it must be largely left to the user as to which analysis is the most appropriate for a particular purpose. The following is a brief comment on each form of analysis.

With subjective ratings users are not limited to making mechanistic counts. Theirs can be a global judgment based on accumulated experience of children and of texts, and taking into account a variety of factors thought to contribute to reading difficulty or ease. If a number of judges is available, then an average judgment is clearly more reliable than any single judgment. The difficulty with subjective ratings, of course, is that a group of judges is not always available; judges are not always conscious of which factors they take into account; and ratings across judges can sometimes vary quite widely. In the present exercise with the three parallel texts of Aesop's fable, while there was general consensus that Passage B was the easiest to read, there was less agreement about the relative reading difficulty of Passages A and C, though most were inclined to rate Passage C as harder than Passage A.

The cohesion analysis provides additional information about texts that may sometimes be useful in explaining unexpected sources of reading difficulty. For many teachers the concpet of cohesion is relatively new. Intuitively, the notion of semantic links without which there would not be text but rather a jumble of sentences, is an attractive one. Much more work needs to be done

in this area though an important start has been made (see, for example, the March 1983 issue of the Australian Journal of Reading which had as its special focus cohesion and the reading teacher). The analyses reported above, where percentage of grammatical and lexical ties, and ties per word and per sentence were calculated must be treated with caution. Insufficient is known about the relationship of cohesion and comprehensibility and although certain trends emerged with the three parallel texts, these must be thought of only as very tentative hypotheses to be explored in the analysis of further texts. Hasan (in press) takes this kind of analysis much further in her examination of central and relevant tokens and the interaction of chains of cohesive ties.

The analysis of selected linguistic variables, as described above, is a well-trodden path and, although the particular instruments described here (Lix and Rix) are relatively new, they are based on a research paradigm going back more than 50 years. The measures are quite objective, by which is meant different analysts would arrive at the same results, as indeed would a computer. However, just because Lix and Rix come up with quantifiable indices, there may be a danger of regarding these as more precise than they really are. Such indices may be useful in providing estimates of difficulty but the user must always remember the underlying assumptions on which such estimates are based, namely that sentence length and word length are themselves substitutes for sentence complexity and word complexity, variables very difficult to measure more precisely and directly.

Finally, all the analyses discussed above are indirect measures for none takes the reader directly into account. It could be argued that such indirect measures therefore lack validity. This is an issue about which there will always be considerable controversy. The view taken here is that, indirect measures though they may be, they may nevertheless provide the teacher with useful insights in critically evaluating texts for use in the classroom.

References

Anderson, J. Cue systems, cohesion and comprehending. Australian Journal of Remedial Education. 14, 56-59, 1982.

Anderson, J. Lix and Rix: variations on a little-known readability index. Journal of Reading. 26, 490-96, 1983.

Australian Journal of Reading (Special Issue - Cohesion and the Reading Teacher), 6, 1-48, 1983.

Halliday, M.A.K. and Hasan, R. Cohesion in English. London: Longman, 1976.

Halliday, M.A.K. and Hasan, R. Text and Context: Aspects of Language in a Social-semiotic Perspective (Sophia Linguistica vi). Tokyo: Graduate School of Languages and Linguistics, Sophia University.

Hasan, R. Coherence and cohesive harmony. In Flood, J. (Ed.) Understanding Reading Comprehension (in press).

Moe, A.J. Cohesion, coherence, and the comprehension of text. Journal of Reading. 30, 16-20, 1979.

Tierney, R.J. and Mosenthal, J. Discourse Comprehension and Production: Analysing Text Structure and Cohesion (Technical Report No. 152) Urbana: Center for the Study of Reading, University of Illinois, 1980.

Passage A: The Dog and His Shadow

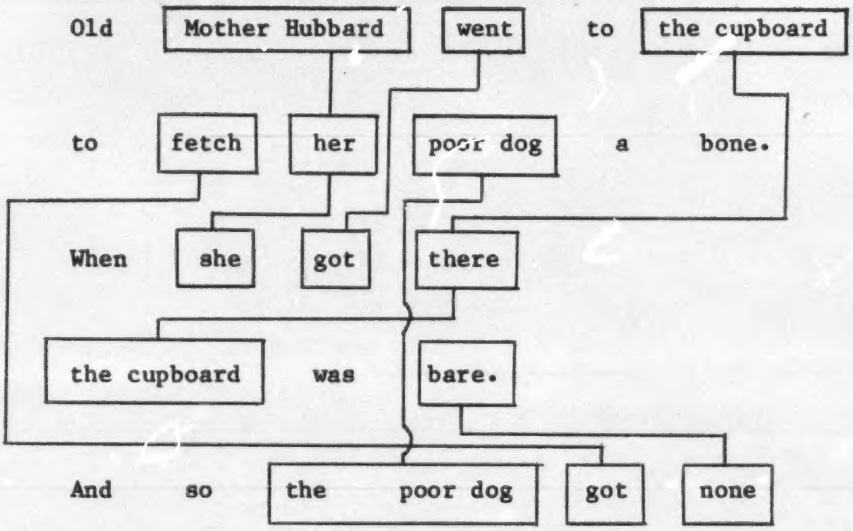
It happened that a dog had got a piece of meat and was carrying it home in his mouth to eat. Now on his way home he had to cross a plank lying across a running brook. As he crossed, he looked down and saw his own shadow reflected in the water beneath. Thinking it was another dog with another piece of meat, he made up his mind to have that also. So he made a snap at the shadow, but as he opened his mouth the piece of meat fell out, dropped into the water, and was never seen more.

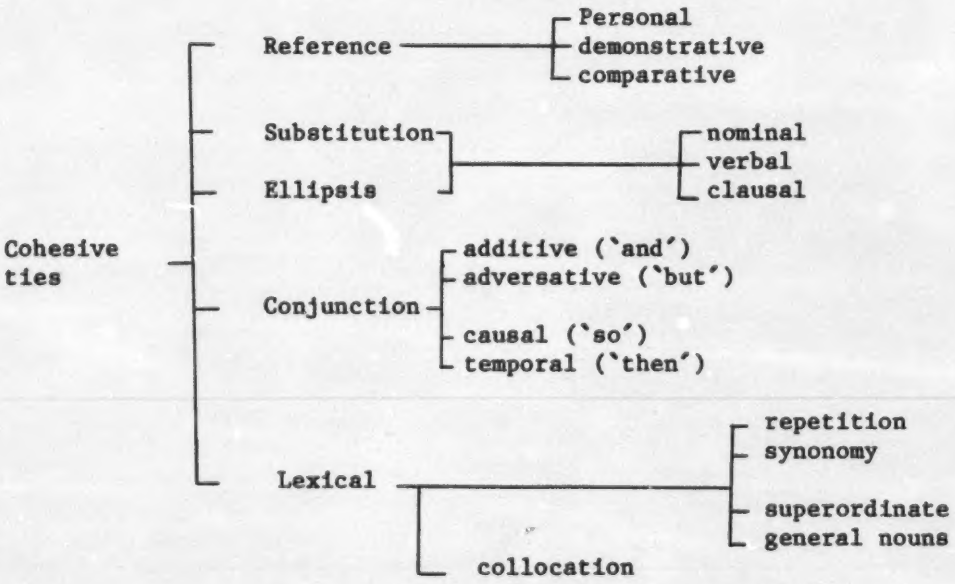
Passage B: The Dog and His Bone

A little dog hurried to the stream with a large juicy bone in his mouth. He wanted to eat the bone all by himself. So he ran across a log that bridged the stream. Then, in the water, he saw a picture of himself. But he thought it was another dog. "Ah, now I shall have two nice bones to eat", thought the greedy little dog. He growled and snapped at the other bone. SPLASH! His bone fell in the water. And so did he! Now he had nothing to eat. Greedy-greedy makes a hungry puppy.

Passage C: The Dog and the Shadow

A dog, crossing a bridge over a stream with a piece of flesh in his mouth saw his own shadow in the water, and took it for another dog, with a piece of meat double his own in size. He therefore let go his own, and fiercely attacked the other dog, to get his larger piece from him. He thus lost both - that which he grasped at in the water, because it was a shadow; and his own because the stream swept it away.





(after Halliday and Hassan 1976)

TABLE 1: Cohesive Ties

	Passage A	Passage B	Passage C
No. of grammatical ties	25	29	19
No. of lexical ties	10	19	12
Total no. of ties	35	48	31
% of grammatical ties	71	60	61
% of lexical ties	29	40	39
No. of words	101	96	84
No. of sentences	5	12	4
Ties per word	.35	.50	.37
Ties per sentence	7.00	4.00	7.75

TABLE 2: Selected Linguistic Variables

	Passage A	Passage B	Passage C
No. of words	101	96	84
No. of long words	10	12	8
No. of syllables	120	116	103
No. of monosyllables	84	79	68
No. of sentences	5	12	4
Sent. length (words)	20.2	8.0	21.0
% long words	9.9	12.5	9.5

ESTIMATING READING DIFFICULTY WITH RIX*

Developed by Jonathan Anderson (Flinders University of South Australia)

Directions

1. Select a sample of sentences from the book(s) to be analysed. It is not possible to be precise about the number of samples since this depends in part on the size of the book(s) and in part on the consistency of writing. As a guide, for short texts, ten 10-sentence samples taken regularly through the book may be sufficient; while for longer works, samples of at least twice this size will probably be required. Very short texts may be analysed in their entirety.

2. For each total sample (excluding headings, captions etc.)

   (a) Count the number of sentences
   A sentence is defined generally as a sequence of words terminated by a full-stop, question or exclamation mark, colon or semi-colon. However, in direct speech, sequences like "Where?" he asked. and "Go!" he ordered. count as single sentences.

   (b) Count the number of long words (i.e. words of 7 or more characters after excluding hypens, punctuation marks and brackets)

   A word is defined generally as a sequence of characters bounded by white spaces. Thus numbers like 1,461 and 10.2, hyphenated sequences, abbreviations (e.g. IRA, a.m.), dates such as (1981-1982), and symbols like % count as single words.

3. Determine Rix by dividing the number of long words by the number of sentences and multiplying by 100.

Interpretation

1. To find the equivalent grade level of difficulty for Rix, enter the Table below and locate the corresponding grade.

Rix score	Rix grade level
0-19	1
20-49	2
50-79	3
80-129	4
130-179	5
180-239	6
240-299	7
300-369	8
370-449	9
450-529	10
530-619	11
620-719	12
720+	College

* This page is not copyrighted and may be reproduced for educational purposes with the usual acknowledgement of source.